POSTPRINT
Published in: Natural Language and Linguistic Theory Vol. 28 (2010)
No. 4, pp. 931-952. DOI: 10.1007/s11049-011-9125-5

Spotting, collecting and documenting negative polarity items

Jan-Philipp Soehn · Beata Trawiński · Timm Lichte

Abstract As the nature of negative polarity items (NPIs) and their licensing contexts is still under much debate, a broad empirical basis is an important cornerstone to support further insights in this area of research. The work discussed in this paper is intended as a contribution to realizing this objective. We briefly introduce the phenomenon of NPIs and outline major theories about their licensing and also various licensing contexts before discussing our major topics: Firstly, a corpus-based retrieval method for NPI candidates is described that ranks the candidates according to their distributional dependence on the licensing contexts. Our method extracts single-word candidates and is extended to also capture multi-word candidates. The basic idea for automatically collecting NPI candidates from a large corpus is that an NPI behaves like a kind of collocate to its licensing contexts. Manual inspection and interpretation of the candidate lists identify the actual NPIs. Secondly, an online repository for NPIs and other items that show distributional idiosyncrasies is presented, which offers an empirical database for further (theoretical) research on these items in a sustainable way.

Keywords Polarity items · Corpus-based retrieval · Empirical database · Documentation · XML

J.-P. Soehn
Stauperle 28, 72108 Rottenburg, Germany
e-mail: jp.soehn@gmail.com

B. Trawiński (✉)
Department of Slavonic Studies, University of Vienna, Spitalgasse 2, Hof 3, 1090 Vienna, Austria
e-mail: beata.trawinski@univie.ac.at

T. Lichte
Emmy Noether Project, SFB 441, University of Tübingen, Hausserstr. 11, 72076 Tübingen, Germany
e-mail: timm.lichte@uni-tuebingen.de

1 Introduction

Negative polarity items (NPIs) are the subject of research in syntax, semantics, and pragmatics, started by Klima's survey of negation in English (Klima 1964). The classical example of an NPI is the English indefinite determiner *any*. As demonstrated in (1), a sentence containing *any* and an item denoting negation is grammatical; without the negation the sentence is ungrammatical.[1] Following standard terminology, we will refer to the negation as the *licenser* or *trigger* of the NPI. We will underline NPIs and typeset the licensers in bold face.

(1) a. He has**n't** seen any dodos.
b. *He has seen any dodos.

The detailed description of relevant features of licensing contexts is a controversial research area since NPIs may occur in the scope of negation as well as in a variety of other environments (such as interrogatives, antecedents of conditionals, or modifiers of superlative and universal NPs). Generalizations such as downward-entailingness (DE, cf. Fauconnier 1975; Ladusaw 1980) are focused on in the literature, whereas other triggers which are not as easy to characterize tend to be ignored. Other approaches discuss some additional aspects of NPI-licensing, cf. Krifka (1995), who aims at a semantico-pragmatic explanation of why a given element is only felicitous in NPI-licensing contexts. However, even if there may be parts of a word's lexical semantics or pragmatics which make it sensitive to polar environments, it sometimes seems to be pure idiosyncrasy that a given word is an NPI, whereas a semantically similar word is not negative polar (e.g. German: sonderlich vs. *besonders* 'particularly').

Making headway in this area requires a certain amount of empirical data that is compiled in a systematic way and that furthermore shows that the phenomenon of NPIs is by no means limited to any, even, or minimizers such as lift a finger. For English and Dutch, the inventory of NPIs has been documented fairly well. Hoeksema (2005) for instance presents about 700 Dutch NPIs. For German, the state of documentation is less ideal. There is only one relatively extensive list in Kürschner (1983), which contains 344 items. However, it needs to be pointed out that Kürschner's listing is far from exhaustive and that we are not in complete agreement with Kürschner's introspectively gained selection. In fact, we have some doubts as to the status of about 200 items (58%) with respect to negative polarity. Even if all of his items felicitously occur in negative contexts, most of them are not NPIs in the sense of being forbidden in non-NPI-licensing environments.

The aim of this paper is twofold: First, we show how statistics can be used to automatically acquire a list of NPI candidates from a partially parsed corpus of written German. Yet we do not claim that a validated and exhaustive list of German NPIs can be obtained with our method. The validation of the candidates has been done by hand and—from our point of view—exhaustivity can never be claimed. Second, we describe our collection of NPIs which represents the result of our spotting algorithm

[1] For the case of *any*, we do not discuss the free-choice item, cf. Kadmon and Landman (1993).

and documents the items together with additional information, providing a freely accessible database on NPIs for further research.

The paper is organized as follows: We provide some theoretical background on NPIs in Sect. 2. Some properties of negative polarity items are discussed, the licensing question is addressed and a possible classification of NPIs is indicated. In Sect. 3, we explain our algorithm to spot NPIs in a corpus and its extension for multi-word NPIs. Subsequently, the resulting candidate list is presented. Section 3 also summarizes our discussion so far before we proceed to our description of collecting and documenting NPIs (Sect. 4). Section 5 rounds off this paper with a conclusion. In the Appendix we provide some of the NPIs we were able to spot using the method discussed in Sect. 3.

2 NPIs: theoretical background

2.1 Negative polarity items

Many approaches to NPIs which can be found in the literature tend to concentrate on a subset of NPIs, mostly quantifying expressions such as any or end-of-scale expressions such as a drop. It must be emphasized that polarity items can be found in almost every part of speech. We give examples of verbal NPIs in (2) and (3), an adverbial NPI in (4) and a nominal NPI in (5). It is noteworthy and will become important later on that NPIs can be single-worded as (2) or multi-worded as (3).

(2) a. Es schert ihn **nicht**.
it bothers him not
'He doesn't give a damn about it.'
b. *Es schert ihn.

(3) a. Er hat es **nicht** wahrhaben wollen.
he has it not accept as true want
'He did not want to accept it as true.'
b. *Er hat es wahrhaben wollen.

(4) a. Hans war **nicht** sonderlich zufrieden mit seiner Arbeit.
Hans was not very happy with his work
b. *Hans war sonderlich zufrieden mit seiner Arbeit.

(5) a. Um acht war noch **keine** Menschenseele da.
At eight was yet no men's soul there
'By eight o'clock no one had arrived yet.'
b. *Um acht war noch eine Menschenseele da.

In these examples, the verbal and the adverbial NPI must be in the direct scope of a negative element at the Logical Form (LF), i.e. no non-DE sentential operator may intervene. Nominal NPIs, often used as minimizers, e.g. roter Heller in (6), impose an additional (syntactic) constraint, namely that the licenser c-commands the minimizer (on the surface level).

(6) a. **Niemand** hat auch nur einen roten Heller gespendet.
nobody has even one red heller donated
'Nobody has donated a red cent.'
b. *Auch nur einen roten Heller hat **niemand** gespendet.

Such nominal minimizers are likely to be accompanied by even, the German auch nur or the widely discussed Dutch equivalent ook maar (cf. Zwarts 1998; Hoeksema and Rullmann 2001). For a detailed crosslinguistic discussion of minimizers see Vallduví (1994).

2.2 Licensers and the licensing property

In the following Table 1, items and constructions that license NPIs are listed, illustrated with the NPI ever and some pertinent references. Hoeksema (2010) gives a similar list of licensers.

Importantly, not all NPIs are necessarily licensed in all of these contexts. The distributional pattern of a given NPI may differ from that of another. This leads to the classification of NPIs (see next section).

Table 1 Licensing contexts

Licensing context	Example	References
negative particles and negative quantifiers	*Nobody will ever...*	Ladusaw (1980) van der Wouden (1997)
neg. conjunctions (*without*)	*Without her ever being...*	van der Wouden (1997)
restrictor of universal quantifiers (except *each* in English) and superlatives	*Everyone, who ever...*	Ladusaw (1980) van der Wouden (1997)
DE quantifiers	*Few people ever...*	Linebarger (1987)
only	*Only Pat has ever been to...*	Linebarger (1987) von Fintel (1999)
non-affirmative predicates (*doubt*, *be surprised*)	*I'm surprised that you ever...*	Linebarger (1987) van der Wouden (1997)
negated neg-raising verbs (*believe*)	*I don't believe that you ever...*	Sailer (2006)
questions	*Did you ever...?*	van der Wouden (1997) Giannakidou (1998)
antecedent of conditionals	*If you'd ever...*	Ladusaw (1980) van der Wouden (1997)
comparative *than*-sentence	*...better than I ever expected.*	Hoeksema (1983)
too-comparatives	*...too many than I could ever count.*	Hoeksema (1983)
negative predicates (*unlikely*)	*It's unlikely that he ever...*	van der Wouden (1997)

Ladusaw (1980) took one of the first steps towards an NPI licensing theory in order to explain the licensing properties of the above contexts. He established that NPIs can only occur in downward-entailing contexts, building on an idea from Fauconnier (1975). Seeing the failure of the standard DE theory to subsume all possible licensing contexts, Giannakidou (1997), building on Zwarts (1995), proposes an approach in which *non-veridicality* is the basic property of NPI licensers. However, although her analysis elaborates on some unresolved issues (e.g. questions are not DE but non-veridical), her theory is less restrictive than required. However, her approach is suitable for the Greek kanenas-indefinites, which have a distribution that is essentially broader than that of any or brauchen 'need'. Another attempt to remedy the shortcomings of the standard DE theory is made by von Fintel (1999), who applies a more restricted, presupposition-dependent notion of entailment (Strawson-entailment), which is also used by Condoravdi (2010).

According to the theories proposed in Kadmon and Landman (1993), Krifka (1995), and Chierchia (2006), which focus on the meaning and pragmatics of NPIs, these items have the semantic properties of domain widening and strengthening. They may introduce alternatives to the foreground information which induce an ordering relation of specificity. The NPI itself denotes the most unspecific element on this scale. Thus, NPIs are banned from non-licensing contexts such as affirmative or upward-entailing contexts because they do not contribute anything new or relevant to the background information.

Another set of licensing theories considers pragmatic factors to an even greater extent. For example, de Swart (1998) argues that the (im)possibility of inverse scope configurations in which an NPI precedes its negative licenser can be explained by considering the pragmatic implicatures triggered by the NPI.

2.3 NPI classification

An important contribution to research on NPIs has been made by Zwarts (1998), who introduces a classification of NPIs by means of their licensing requirements. He distinguishes between different, logically defined categories of licensers which exhibit different grades of negativity. We maintain Zwarts' concepts but adopt the terminology in van der Wouden (1997) and differentiate between minimal (downward-entailing expressions, e.g. *few*), regular (anti-additive expressions, e.g. *nobody*) and classical (anti-morphic expressions, e.g. *not*) negation.[2] One can classify NPIs into superstrong NPIs (licensed only by classical negation), strong NPIs (licensed also by regular negation), and weak NPIs (licensed in all three contexts). The NPI classes and their distributional pattern across the grades of negation are depicted in Table 2.

For example, the NPI anders können 'can(not) help doing' is compatible with classical and regular negation, but is excluded from minimal negation: "Ich konnte nicht anders als zu helfen." 'I couldn't help helping.', "Niemand konnte anders als zu helfen." 'Nobody could help helping.' but "*Wenige konnten anders als zu helfen." 'Few could help helping.' Therefore, it can be classified as a strong NPI.

[2]For the exact definitions of downward-entailing, anti-additive, and anti-morphic see Zwarts (1996).

Table 2 NPI classes

NPI	Negation		
	classic	regular	minimal
weak	+	+	+
strong	+	+	−
superstrong	+	−	−

The expression missen wollen 'to (not) want to do without' is compatible with all kinds of negation, which makes it a weak NPI: "Die Einparkhilfe will ich nicht missen!" 'I don't want to do without the parking assistant!', "Die Einparkhilfe will niemand/wollen nur wenige missen, wenn sie einmal in den Genuss gekommen sind." 'No one/only few people would want to do without a parking assistant, once they benefited from this system in their car.'

Expressions classified as superstrong are mostly idiomatic, e.g. sich so haben in "Hab dich nicht so!" 'Don't get your knickers in a twist!'. Thus, we entertain some doubt that there are plain superstrong NPIs in German.

3 Spotting NPIs in a corpus

The basic idea behind the corpus-based retrieval mechanism described here is to regard the relation between an NPI and its licenser as being similar to the relation between a collocate and its collocator. This idea, going back to van der Wouden (1992) and then pursued in van der Wouden (1997), allows us to apply regular collocation acquisition techniques in order to obtain a list of NPI candidates. The aim of our efforts is not to validate items that are assumed to be negative polar. Rather, we aim for a list of NPI-candidates, i.e. a rich source for collecting NPIs. In other words, the resource we are providing is meant to serve as input for the "NPI-seeking" linguist.

3.1 The algorithm

The procedure of NPI extraction is divided into three steps: conversion of the corpus, lemma counting and evaluation of frequency data, resulting in a lemma ranking.

3.1.1 Corpus conversion

The corpus on which the extraction algorithm is performed is a part of TüPP-D/Z (*Tübingen Partially Parsed Corpus of Written German*).[3] TüPP-D/Z consists of about 200 million words, based on the electronic version of the German newspaper *die tageszeitung* (*taz*). It contains lemmatization, part-of-speech tagging, chunking and clause boundaries. The section of TüPP-D/Z that we used in this study consisted of about 5.8 million sentences from the years 1990 to 1998. Given the fact that many NPIs are sparse, the size of the underlying corpus is of great importance. In addition,

[3] Cf. www.sfs.uni-tuebingen.de/tuepp.shtml.

the provided annotation is crucial with respect to the identification of the licensers and their scope.

The first step was the conversion of the corpus so that it contained only lemmatized words and the clause structure. Concurrently, licensers were identified and annotated with the aid of POS tags and chunking. The corpus, then, consisted of sentence strings such as the following:

(7) <CL1> von Friede können also bei alle Optimismus noch
of peace can therefore at all optimism yet
lange DEINT die Rede sein. </CL1>
long DEINT the talk be
'Even with all optimism, we won't yet be able to talk about peace for a long time.'

<CL1> and </CL1> represent the clause boundaries, and the licenser *nicht* 'not' is replaced by the licensing marker DEINT ('downward entailing and interrogative'), which is the marking for a licenser.

Unfortunately, some of the licensers mentioned in Table 1 cannot be clearly identified in the corpus, if at all. For this reason, we tried to avoid ambiguous cases and preferred less licenser annotation rather than risking incorrect annotation. Nevertheless, NPIs should still show a significant distributional pattern, while statistical noise should be suppressed. These hidden licensers are: covert conditionals, extraposed relative clauses as restrictors of superlatives and universal quantifiers, comparative than-clauses.

As Hoeksema (1997) points out, conditionals can be covertly realized, as in: *You say anything, and I'll kill you*. We think that it is virtually impossible to reliably identify these licensers with current corpus annotation.

Restrictors of superlatives and universal quantifiers are only marked up if they are relative clauses. If these relative clauses are moved to the right periphery of the sentence (into the so-called *Nachfeld*) they cannot be related to the superlative and universal quantifier, respectively, in a direct way. Instead of annotating all relative clauses in the Nachfeld of a sentence where these licensers occur, we dropped cases where the relative clause did not immediately follow the superlative and universal quantifier, respectively. Analogously, *than*-clauses in comparatives were ignored since they mostly occurred in an extraposed position. Furthermore, the corresponding comparative constructions are hard to detect in general. Needless to say, it is unavoidable to have gaps in lexically open licenser classes such as non-affirmative verbs, negated neg-raising verbs and negative predicates.

3.1.2 Counting of lemmas

After corpus conversion we extracted for each lemma in the corpus the number of total occurrences and the number of occurrences in clauses which contained a licensing marker (DEINT). The somewhat simplified scope of a licensing marker was taken to be exactly the clause in which it occurred. It did not comprise embedded clauses.[4]

[4]We ignore cases of double negation, being aware that this potentially leads to additional statistical noise. In fact, about 10% of the licensing clauses contain at least two licensing markers, but one has to take into

Note that in cases where an item is able to license NPIs in an embedded sentence, e.g. with negated neg-raising verbs or inherently negative verbs, the licensing marker was added to the embedded sentence.

3.1.3 Evaluation of frequency data

In order to obtain a list of NPI candidates, we compiled a lemma ranking based on a very straightforward association measure that we will call the *context ratio* (CR). It is computed for a lemma l using its overall frequency N and the frequency N_{lic} of configurations where l is in the scope of a licensing marker:

(8) $$\text{CR} := \frac{N_{lic}}{N}$$

CR reflects the fraction of the licensed occurrence of a lemma relative to its overall occurrence, and hence it ranges between 0 and 1. We expect NPIs to have a significantly high CR value, but not necessarily 1. CR does not differ from association measures such as *pointwise mutual information* (MI, Church and Hanks 1990) in our setting. The reason for this is the semi-fixed nature of our bigrams and the fact that we are not interested in the actual values of an association measure, but in the broad lemma ranking based on it. Given a lemma w with frequency N_w, the frequency of negative contexts N_{neg} and furthermore $N_{w,neg}$ as the frequency of w occurring in a negative context, the formal definitions of CR and MI will then appear as follows:

(9) $$\text{CR} := \frac{N_{w,neg}}{N_w}$$
$$\text{MI} := \frac{P(w,neg)}{P(w)*P(neg)} = \frac{N_{w,neg}/N}{(N_w/N)*(N_{neg}/N)} = \frac{N_{w,neg}}{N_w} * \frac{N}{N_{neg}}$$

$P(w, neg)$ is the probability of the co-occurrence of w and a negative context. It is obvious that $\frac{N}{N_{neg}}$ has a constant value and hence is not substantial for the computation of the ranking.

Due to the weakness of MI and CR against low frequencies, we integrated a cut-off at a frequency of 30, so that the size of the lemma list was reduced to 64,867 lemmas (from 1,058,462 lemmas). To avoid this trade-off one could consider using Pearson's chi-squared test or Log-likelihood (cf. Manning and Schütze 1999) as association measures. However, these are biased in favor of lemmas with high frequency counts (cf. Evert 2005: 84). In order to get reasonable rankings, we scaled association measures based on them by the overall frequency, finally ending up with rankings almost congruent with those of the much simpler MI and CR. Therefore we adhered to the latter measure.

3.1.4 Enhancement for multi-word NPIs

Only single lemmas have been considered so far. However, we know that many NPIs are complex and only show negative polarity as complex entity, i.e. the combined

account that, for example, wh-questions receive two licensing markings for both the wh-pronoun and the question mark.

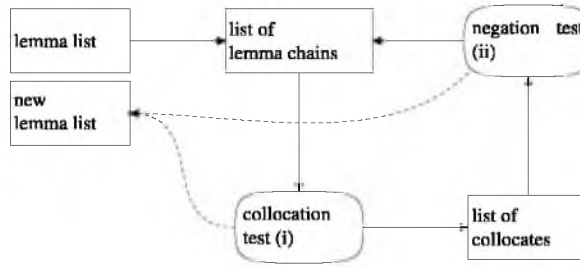

Fig. 1 Schema of the enhancement for the retrieval of complex NPIs

lemmas are not inherently negative polar, but the combination of the lemmas is. For example, the noun *Tassen* 'cups' is not an NPI and ranges very low on a preliminary candidate list. The idiom (nicht) alle Tassen im Schrank haben 'to have lost one's marbles' is an NPI. To spot it in the corpus, we need to find some other parts of that expression. Therefore, we enhanced the extraction algorithm to also include complex NPIs. A schema of the new algorithm is depicted in Fig. 1.

The starting point of our enhanced algorithm is the list of lemmas and their context ratios. We performed a *collocation test* for every lemma to identify other lemmas that significantly co-occur (i) in the same clause and (ii) in negative contexts. As a collocation measure we integrated the G^2 score from Rayson and Garside (2000), a simplified Log-likelihood calculation. By rule of thumb, we let those collocates pass that had a G^2 value of $\geq$250 and still co-occurred more often than a certain minimal frequency threshold ($N \geq 10$). This yielded a list of collocates for each of the lemmas.

Next we checked whether the distribution pattern of lemma and collocate showed higher or equal affinity for negative contexts than the lemma individually (*negation test*). If that was the case, we repeated the procedure on the lemma-collocate pair, which was now handled the way we handled single lemmas. This resulted in *chains of lemmas* as new NPI candidates, which could eventually not be expanded further because they lacked either collocates or an enlarged affinity for negation. These new lemma chains were added to the original lemma ranking in accordance with their context ratio. Starting for instance with the lemma *Tasse* 'cup' ranked at 15,221, the enhanced acquisition method compiles the lemma chain *Tasse Schrank* 'cup cupboard' which corresponds to the negative-polar expression alle Tassen im Schrank haben and to position 433 of the ranking.

Thus, the enhancement did not only generate lemma chains that are easier to map onto complex NPI candidates, but it also moved those complex NPIs whose parts are rather non-polar and hidden at lower rankings to a more prominent position on the list.

3.2 Result: the candidate list

By performing the method described in the previous section, a ranking consisting of single lemmas and lemma chains was generated. The linguist who is interested in collecting NPIs will go through these by hand, expecting good NPI candidates to have accumulated among the higher ranked entries. Since the ranking comprises about 65,000 items, the question may arise as to how far one should go down. This decision

rests with the researcher. In the following case, however, we singled out a shortlist of NPI candidates from the top of the overall ranking in order to evaluate the retrieval method. The shortlist was derived with the aid of *z-values* (Moore and McCabe 2006) that standardized the CR values and permitted us to determine significantly deviating CR values.[5] In our setting, the derived shortlist comprised about 2000 single lemmas and lemma chains (at $p \leq 0.01$). The evaluation was done by a linguist who inspected the lemmas and used corpus evidence to support a possible NPI-hood.

Among the 50 lemma chains with the highest scores, most of the items can indeed be connected to negative polarity. For example, one finds lemma chains such as in (10) that can be mapped onto the complex NPIs etw. unversucht lassen, aus dem Staunen herauskommen, and vorne und hinten reichen:

(10) a. `unversucht lassen`
untried let
'to leave sth. undone'
b. `Staunen heraus`
amazement out
'being constantly amazed' (with negation)
c. `reichen vorne hinten`
suffice in front behind
'to be totally sufficient'

Interestingly, some NPIs appear to be somewhat hidden in different lemma chains. For example, *brauchen* 'need', which is known to be negative polar when used as an auxiliary verb, can only be found as a part of five lemma chains. In this case, however, it does not seem to be a serious problem, since these lemma chains reflect regular complex expressions that allow the recovery of the negative polar *brauchen*. Something similar can be observed for the intensifiers *gar* 'at all', and *mehr* 'more' and for the particle *noch* '[not] yet'. The rank of some NPIs often discussed in the literature is shown in (11):

(11) 268: *eine Menschenseele* 'a soul'
284: *sonderlich* 'particularly'
449: *einen Hehl machen aus* 'to make no secret of something'
646: *sich scheren um* 'to give a damn about something'
784: *jemals* 'ever'

One has to carefully distinguish complex NPIs from idiomatic expressions that include a *certain* negative element. In our list we found e.g. the corresponding lemma chain for the advertising slogan *Man gönnt sich ja sonst nichts* (meaning: 'It's my one and only treat.') with a rather fixed negative word (*nichts*) in the object position.

Yet not all of the lemma chains are connected to negative polarity per se, i.e. many show an affinity for negation that is triggered by the style of the newspaper text. Nevertheless these items can still occur outside licensed contexts and therefore

[5]Note that z-values are intended for data sets with normal distribution. However, the CR values of our single lemmas are skewed towards a CR score of 1, the mean being at 0.1178. Since we only use z-values for deriving a shortlist, we ignore this.

they are called *pseudo NPIs* following Hoeksema (1997). In (12) we give the lemma chains for *hinter dem Ofen hervorlocken*, *einer gewissen Komik entbehren*, and *bei Redaktionsschluss noch (nicht) feststehen*:

(12) a. `hinter Ofen hervor locken`
behind oven out tickle
'to get someone excited about something'
b. `entbehren gewiß Komik`
lack certain humor
'to be lacking in humor'
c. `Redaktionsschluss fest stehen noch bei`
press date fixed be yet at
'be certain at press date'

Finally, we also found seven instances of statistical noise, i.e. lemma chains which undoubtedly have nothing to do with negative polarity or which the authors fail to map on any reasonable complex expression. They can be suppressed quite easily because of their CR value of 1 and their noticeable length. In (13) we present such a lemma chain that arises due to a recurring statement in the weekly 'Letters to the editor' section of the corpus newspaper (*Die auf dieser Seite erscheinenden Leserbriefe geben nicht notwendigerweise die Meinung der taz wieder.*).

(13) `notwendigerweise Meinung Seite erscheinend`
necessarily opinion page figuring
`geben wieder auf die`
reflect on the
'The readers' letters on this page don't necessarily reflect the opinion of the taz.'

Only a very small part of the shortlist has lemma chains with a CR value of 1, just as the theory predicts for NPIs. Yet in reality, we have to accept the following inevitable shortcomings: firstly, we simply do not know all of the possible licensers. Secondly, as mentioned in Sect. 3.1.1, not all licensers can be unambiguously identified and annotated in the corpus. Thirdly, polysemy is virtually omnipresent among polarity items; that is, many NPIs have non-polar counterparts that leave their marks on the CR value.[6] Fourthly, also the applied frequency cut-off certainly has an impact on the emergence of false negatives.

3.3 Interim summary

We proposed a method for spotting NPI candidates in a partially parsed corpus. The core idea was to obtain for every lemma of the corpus the context ratio with respect

[6]For example, the verb *angehen* has at least four readings: (1) *das Licht geht an* 'the light is turned on', (2) *wir gehen das folgendermaßen an* 'we'll tackle this in the following way', (3) *es geht dich nichts an* 'it's none of your business' (4) *es kann nicht angehen, dass* ... 'it can't be true that ...'. Only (3) and (4) can be considered as negative polar. A way to discriminate between these different meanings could be to integrate automatic word-sense disambiguation, see also Sect. 5.

to licensing expressions and to derive a ranking based on these context ratios. As many NPIs consist of more than one word, we enhanced our algorithm to extract not only simple words but also complex expressions. In line with our expectations, we found promising NPI candidates among the topmost lemmas of this ranking. Note that this method can be expanded to also subcategorize NPIs along Table 2 using their statistical profiles (Lichte and Soehn 2007).

The expressiveness of our method with respect to obtaining cast-iron NPIs is rather restricted. This is mainly due to the nature of a text corpus and due to the difficulty in correctly identifying and annotating licensers. There still remains a considerable amount of work for the linguist as far as the evaluation of candidates and the compilation of a comprehensive list of NPIs for a given language is concerned.

4 Collecting and documenting NPIs

The application of the algorithm described in Sect. 3 resulted in a collection of NPIs, integrated into a bigger linguistic resource on lexical items with idiosyncratic occurrence patterns. This resource is called *The Collection of Distributionally Idiosyncratic Items*, abreviated as CoDII. In Sect. 4.1, we describe the general idea behind this resource, explain its conceptual design and present the currently available subcollections. Sections 4.2 and 4.3 are devoted to the modeling of NPIs in CoDII and to the technical realization of our resource as an online repository. Visualization and quering of CoDII is discussed in Sect. 4.4. In Sect. 4.5, we sum up the presentation of CoDII and indicate how our resource might be further developed in the future.

4.1 CoDII

The essential idea behind CoDII is to provide an empirical basis for linguistic investigations into lexical items showing distributional idiosyncrasies. This includes collecting and listing the particular items, providing existing linguistic documentation, and specifying possible ways of extracting data related to these items. The conceptual design and the data structure of CoDII have been conceived in such a way that on the one hand, subcollections of various types of distributionally idiosyncratic items can be modeled, such as anaphors, negative and positive polarity items, or bound words, and, on the other hand, collections of distributionally idiosyncratic items from various languages can be compiled. Currently, five collections of distributionally idiosyncratic items are available in CoDII: Bound Words of German (CoDII-BW.de), Bound Words of English (CoDII-BW.en), Positive Polarity Items (PPIs) of German (CoDII-PPI.de), NPIs of Romanian (CoDII-NPI.ro), and NPIs of German (CoDII-NPI.de). All collections are available online at www.sfb441.uni-tuebingen.de/a5/codii.

At present, there are two collections of NPIs available in CoDII: CoDII-NPI.ro and CoDII-NPI.de. CoDII-NPI.ro currently includes about 60 Romanian NPIs. The items of CoDII-NPI.ro correspond to the English, German and Dutch NPIs discussed in linguistic literature, since there is no specific collection of Romanian NPIs available. CoDII-NPI.de contains German NPIs and currently includes over 100 items. At

present, adverbial, prepositional, nominal and verbal NPIs are collected and documented. The sources used for acquiring the NPIs for CoDII-NPI.de include the collections of NPIs in Welte (1978) and Kürschner (1983), as well as the list of the NPI candidates automatically extracted from the *Tübingen Partially Parsed Corpus of Written German* (*TüPP*) as described in the previous section. In the following paragraphs, we will describe how the particular NPIs are integrated in CoDII.

4.2 Modeling NPIs in CoDII

Each NPI is characterized in CoDII by four information blocks: General Information, Syntactic Information, Licensing Contexts and Class.

The block General Information identifies an NPI by providing the word form, the English gloss for the (German or Romanian) NPI, the English translation of the (German or Romanian) NPI, possible expressions in which the NPI occurs, and, if appropriate, possible paraphrases of these expressions. In the case of polysemic expressions, only their negative-polar readings are considered. Their non-negative-polar readings are ignored in our resource.

Within the block Syntactic Information, details on the syntactic category of an NPI are given and, if appropriate, the syntactic structure of the expression in which the NPI occurs. The Stuttgart-Tübingen Tagset (STTS)[7] has been used for the syntactic description of German NPIs and expressions in which they occur. For the syntactic description of Romanian NPIs we took the (modified) tagset from the Multilingual Text Tools and Corpora for Central and Eastern European Languages (MULTEXT-East).[8]

The block Licensing Contexts provides information on the licensing environment of a given NPI. The following licensing contexts are taken into account:

- Clausemate negation (cmn): There is a negation particle (*not*) in the same clause.
- Non-clausemate negation (ncnm): Negation particle (*not*) occurs in the matrix clause, while the NPI appears in the subordinate clause.
- Negative particle/negative quantifier (nw): NPI is in the scope of a negative particle or a negative quantifier such as *nobody*, *nothing*, *never* or in their equivalent terms in the other languages.
- '*kein*-negation' (kein): NPI includes *kein* (for German).
- *without*: NPI is in the scope of *without*.
- Universal quantifier (univ): NPI is in the restrictor of a universal quantifier.
- Other downward-entailing contexts (dent): NPI is in the scope of a downward-entailing expression such as *few* or *hardly*.
- *only*: NPI is in the scope of *only*.
- Negative verb (nv): NPI is in the scope of a non-factive predicate such as *doubt*, *fear*, or *it is impossible/improbable that* or of an adversative attitude predicate such as *be surprised* or *regret*.
- Question (que): NPI occurs within a question.

[7] http://www.sfs.uni-tuebingen.de/Elwis/stts/stts.html.

[8] http://nl.ijs.si/ME.

- Conditional (if): NPI is in the restrictor of a conditional operator such as *if*.
- Comparative (comp): NPI is in the *than*-clause of a comparative.
- Superlative (sup): NPI is in the restrictor of a superlative.
- Imperative (imp): NPI occurs within an imperative clause.

In addition, exceptional cases can be specified, i.e. corpus evidence for an NPI that does not occur in a licensing context.

For each licensing context, corresponding examples are provided. The examples for the Romanian NPIs have been acquired from the Romanian electronic corpus developed by Rada Mihalcea from the Department of Computer Science and Engineering at the University of North Texas, USA, from the Romanian electronic corpus developed at the Romanian Academy Center for Artificial Intelligence (RACAI), as well as from the internet via Google. A number of examples have been constructed by Gianina Iordăchioaia, a native speaker of Romanian who collaborated on CoDII-NPI.ro. To acquire the examples for the German NPIs, corpora of the Institute of German Language in Mannheim[9] and the Internet via Google have been used.

Finally, the information block Class specifies the type of polarity, which is always negative in the case of NPIs, and the class associated with a given polarity item. In CoDII, we use the following classes of NPIs with the following definitions:

- Superstrong: NPIs are superstrong if they are licensed only by antimorphic contexts (overt negation).
- Strong: NPIs are strong if they are licensed by antimorphic and anti-additive contexts.
- Weak: NPIs are weak if they are licensed by antimorphic, anti-additive, and downward-entailing contexts (plus the remaining ones).
- Open: for undefined classification.

In the next section, we will illustrate how this conceptual design is realized technically.

4.3 Representing NPIs in CoDII

CoDII-NPI.ro and CoDII-NPI.de have been internally encoded in the Extensible Markup Language (XML). The Document Type Definition (DTD) for CoDII has been specified in such a way that the element `codii` constitutes the document root and its instance is identified by the attribute `type` (for specifying the collection type) and the attribute `xml:lang` (for specifying the language of the data). The content model of the element `codii` consists of two elements: `dii-list`, whose content is a list of NPIs, and `dii-examples`, whose content is a list of examples.

The content model of the element `dii-list` is a list of `dii-entry` elements, which consists of a set of elements that identify NPIs (`dii`), describe documentation on each NPI (`dii-classification`), present syntactic properties of NPIs (`dii-syntax`), and specify licensing contexts of NPIs (`licensers`). Figure 2 presents a fragment of the CoDII-XML-encoding of the German NPI beileibe 'by

[9] http://www.ids-mannheim.de/cosmas2/.

```
<dii-entry id="beileibe">
    <dii>
        <ol>beileibe</ol>
        <en>by no means / at all</en>
    </dii>
    <dii-classification>
        <dii-class category="pi" subcategory="npi"
           type="own" class="weak"
           original-class="no">
           <bibliography bib-item="own"> </bibliography>
        </dii-class>
        <dii-class category="pi" subcategory="npi"
           type="Kuerschner83" class="OPEN"
           original-class="no">
           <bibliography bib-item="Kuerschner:83">
           </bibliography>
        </dii-class>
    </dii-classification>
    <dii-syntax
        hits="beileibe-01 beileibe-02 beileibe-03
              beileibe-04 beileibe-05 beileibe-06
              beileibe-07" cat="ADV">
        <dii-expression-syntax>ADV
        </dii-expression-syntax>
    </dii-syntax>
    <licensers>
        <cmn given="yes" hits="beileibe-01"/>
        <ncmn given="no"/>
        <kein given="yes" hits="beileibe-02"/>
        <nw given="yes" hits="beileibe-03"/>
        <dent given="yes" hits="beileibe-04"/>
        <nv given="no"/>
        <que given="no"/>
        <imp given="no"/>
        <if given="no"/>
        <without given="no"/>
        <only given="yes" hits="beileibe-05"/>
        <univ given="no"/>
        <comp given="yes" hits="beileibe-06"/>
        <sup given="yes" hits="beileibe-07"/>
        <exc given="no"/>
    </licensers>
</dii-entry>
```

Fig. 2 The CoDII-XML-representation of the German NPI beileibe 'by no means'

no means' according to this representation model. This item belongs to a subclass of NPIs corresponding to modifiers that intensify negative particles or negative quantifiers, similar to at all in English. Other examples are gar and überhaupt.

The XML-representation in Fig. 2 demonstrates that the NPI beileibe 'by no means' is identified in CoDII via the attribute `id` and is characterized by the general information (`dii`) including the specification of this NPI in its original language (`ol`) and its English translation (`en`). Further, the information on the available classifications of this NPI is provided (`dii-classification`). Thereby, two classes (`dii-class`) are specified for beileibe according to two sources or classifications.

These include our own, project-internal classification (`type="own"`) and the classification in Kürschner (1983) (`type="Kuerschner:83"`). In both classifications, beileibe is considered a polarity item (`category="pi"`) of the negative polarity type (`subcategory="npi"`). This NPI is defined as weak according to our project-internal classification, based on the logical properties of the licensing contexts (`class="weak"`), and it is assigned the class open (`class="OPEN"`), as associated with the collection of NPIs in Kürschner (1983). The information on the available classifications also includes bibliographic references (`bibliography`), linked via the attribute `bib-item` with the corresponding bibliographic item(s) listed in the electronic collection of publications concerned with NPIs. Further, the NPI beileibe is characterized syntactically (`dii-syntax`). It is assigned the syntactic category of adverb (`cat="ADV"`) and is identified as a single word expression (`dii-expression-syntax`). Finally, a set of licensing contexts is provided. For each context, it is specified whether or not this context licenses beileibe (`given="yes"` vs. `given="no"`). According to the specifications in Fig. 2, this NPI is licensed by clausemate negation, 'kein'-negation, negative particles and negative quantifiers, downward-entailing contexts, by *nur* 'only', comparatives and superlatives. For each context which licenses this NPI, a set of examples from the Internet or the corpora of the Institute of German Language in Mannheim (access via COSMAS II) is provided (`hits`).

All representations of examples associated with the NPIs in CoDII are contained in the same document as the NPIs themselves, and, as already indicated, are encoded by means of the element `dii-examples`. The content model of the element `dii-examples` consists of a list of the `example` elements. The `example` elements are linked to the appropriate NPIs by the attributes `dii` and are identified via the attribute `id`.

4.4 Visualizing and searching for NPIs in CoDII

CoDII-NPI.ro and CoDII-NPI.de are available on the Internet at www.sfb441.uni-tuebingen.de/a5/codii in the form of a set of XHTML files generated by an XSLT script. The CoDII-XML-description of the NPI beileibe in Fig. 2 and the representation of the examples for this NPI are visualized in CoDII as shown in Fig. 3. All tags, licensing contexts, bibliographic specifications and classes are linked with the corresponding information sites.

CoDII not only collects, documents and (alphabetically) lists NPIs, but also offers database functionalities. Due to the integration into an XML database,[10] quering CoDII is possible with respect to the particular lemma, syntactic properties, classifications and classes, and licensing contexts.

4.5 2nd Interim summary and outlook

In this section, we have introduced two electronic collections of NPIs, CoDII-NPI.ro and CoDII-NPI.de and presented how NPIs are modeled, represented, visualized and

[10] We use the Open Source XML database eXist, available at http://exist.sourceforge.net/.

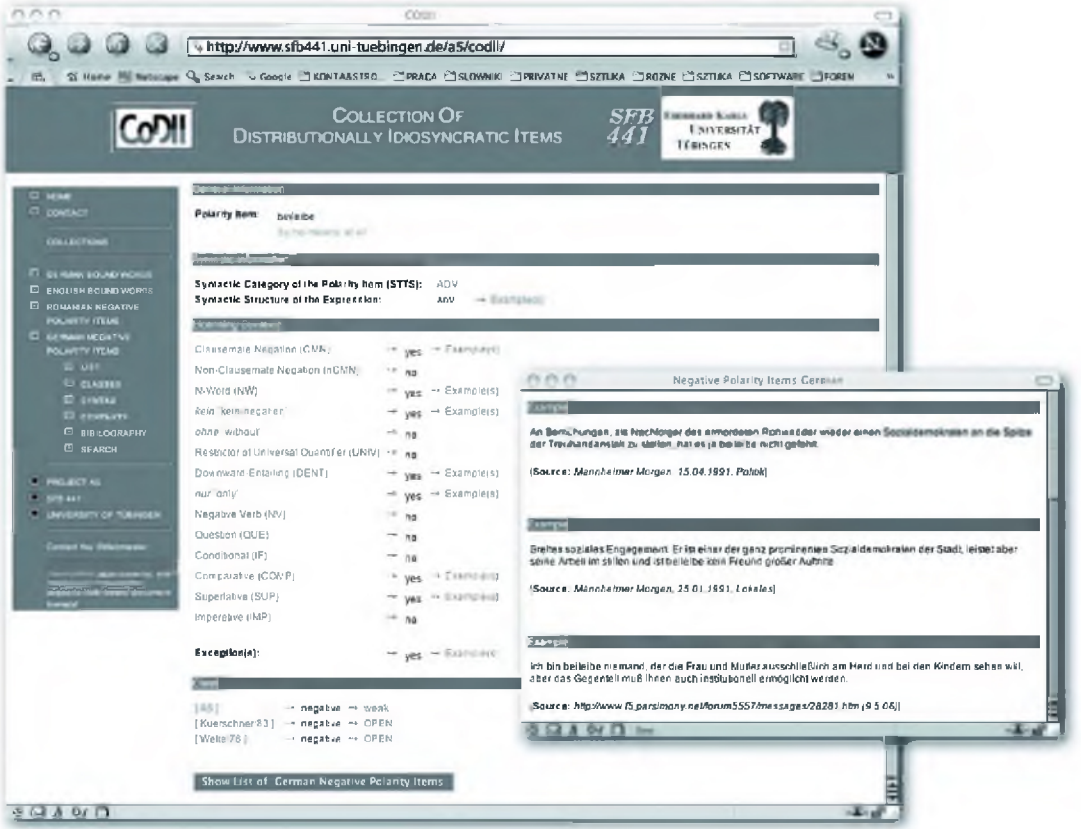

Fig. 3 The browser display for the NPI beileibe 'by no means' and its examples

searched for in CoDII. The architecture of the data structure of CoDII-NPI.ro, and CoDII-NPI.de as well as the remaining collections in CoDII is uniform and linguistically motivated. Its design makes it possible to add further classifications, corpora and search tools, as well as further collections of distributionally idiosyncratic items. An extension to other languages is possible as well.

5 Conclusion

Answering the question of why certain lexical elements are sensitive to certain kind of contexts is an interesting field of research. Many theories exist that try to explain NPI-hood in different ways. Quantifiers such as any have attracted a lot of attention and the fact that there is a free-choice item *any* in English has led to theories that try to combine the two properties of being free-choice and NPI (cf. Kadmon and Landman 1993; Chierchia 2006). However, a solid database for research and theory development of NPIs, e.g. for German, has remained a desideratum so far. We hope to contribute to the lively field of research by our collections of NPIs, as they are conceived as an electronically accessible research platform. CoDII-NPI reveals the great variety that exists among NPIs and the obvious question is whether to treat the different parts-of-speech, idioms and non-idioms, quantifiers, verbs, etc.—all being NPIs—in a uniform way. Discovering subclasses and (re)categorizing NPIs is work that still needs to be done. Detailed corpus studies and psycholinguistic experiments (cf. e.g. Drenhaus et al. 2004; Saddy et al. 2004; Vasishth et al. 2006) seem to be very promising ways to get insights about subclasses of NPIs and their interaction. A further refinement of our spotting algorithm may be done by using a corpus which is word-sense tagged or which contains information about different word uses (cf. Agirre and Edmonds 2006: Chap. 1, for an overview of current research activity on word-sense disambiguation). Many NPIs are homonyms to non-NPIs: For example, (i) *tragbar* means both 'bearable' and 'portable', (ii) *ausstehen* means '(cannot) stand so' and 'be outstanding', and (iii) *scheren* means 'care about' and 'shear' (in each case the former being the NPI meaning, the latter the non-NPI meaning). A disambiguated input corpus might be able to promote real NPIs to a more prominent ranking on the candidate list.

Acknowledgements We are grateful to Lothar Lemnitzer, Mingya Liu, Frank Richter, and Manfred Sailer for their help and discussions. We also want to thank the audience of the Workshop on Negation and Polarity and two anonymous reviewers for their suggestions and comments. The research presented in this contribution was funded by *Deutsche Forschungsgemeinschaft* (DFG) and *LANDESSTIFTUNG Baden-Württemberg*.

Appendix

In the following table we provide 30 of the German NPIs we found with the method sketched in Sect. 3. We give the item, a short expression containing the item itself and classical negation, and an English translation of that expression, where we emphasize

the word matching the German NPI.[11] We hope to exemplify the variety of NPIs with regard to part-of-speech, idiomaticity, and multi-wordedness. More information about the items such as their syntactic category and their licensing environments can be found in our online repository, CoDII-NPI.de.

ausstehen	Ich kann sie nicht ausstehen. 'I can't *stand* her.'
beileibe	Er ist beileibe kein Genie. 'He is *by no means* a genius.'
beirren	So leicht ließ er sich nicht beirren. 'He couldn't be *put off* easily.'
brauchen	Du brauchst nicht zu kommen. 'You don't *need* to come.'
mit rechten Dingen zugehen	Das geht doch nicht mit rechten Dingen zu! 'There must be *something odd* here!'
erst gar	Fang mir erst gar nicht davon an! 'Don't *even* mention that to me!'
lange fackeln	Er fackelte nicht lange. 'He didn't *vacillate* for long.'
gefeit	Vor Misserfolg ist man nicht gefeit. 'One isn't *immune* to failure.'
geheuer	Die Höhle ist mir nicht geheuer. 'That cave is a bit *scary*.'
Hehl machen aus	Er macht keinen Hehl aus seiner Abneigung. 'He doesn't *make a secret* out of his antipathy.'
nachstehen	Sie steht ihm in nichts nach. 'Her skills *are* nothing *short of* his'.
lumpen	Lassen Sie sich nicht lumpen! 'Don't be *stingy*!'
so recht	Ich wußte nicht so recht, was ich tun sollte. 'I didn't *quite* know what to do next.'
hinten und vorne reichen	Das Geld reicht hinten und vorne nicht. 'The money is *really* not *enough at all*.'
scheren	Ihr Unglück schert ihn nicht. 'He doesn't *care* about her bad luck.'
jedermanns Sache	Spinat ist nicht jedermanns Sache. 'Spinach is not *everyone's cup of tea*.'
blasser Schimmer	Davon habe ich keinen blassen Schimmer. 'I haven't *the faintest idea* about it.'
sonderlich	Hier ist nicht sonderlich aufgeräumt. 'It is not *particularly* tidy in here.'
aus dem Staunen herauskommen	Er kam aus dem Staunen nicht mehr heraus. 'I was *wondering and wondering*.'
alle Tassen im Schrank haben	Peter hat nicht alle Tassen im Schrank. 'Peter has *lost his marbles*.'
tragbar	Sein Verhalten ist nicht tragbar. 'His behavior is *unbearable*.'
seinen Augen trauen	Sie traute ihren Augen nicht. 'She couldn't *believe her eyes*.'
überhaupt	Er kennt sich überhaupt nicht aus. 'He doesn't know the place *at all*.'
zu unterschätzend	Ein nicht zu unterschätzender Sturm näherte sich der Küste.

[11] Note that the English equivalent is sometimes not the literal translation and is not even necessarily an NPI.

	'A storm is approaching the coast which mustn't be *underestimated*.'
umhinkommen	Ich kam nicht umhin, darüber zu lachen. 'I could't *refrain from* laughing about it.'
unversucht	Wir ließen nichts unversucht. 'We left nothing *untried*.'
verwunderlich	Das ist nicht weiter verwunderlich. 'This is not very *remarkable*.'
von ungefähr	Ihr Erfolg kommt nicht von ungefähr. 'Her success is *no accident*.'
wahrhaben wollen	Sie wollte seinen Tod nicht wahrhaben. 'She refused to believe that he was dead.'
wegdenken	Er ist aus dem Chor nicht mehr wegzudenken. 'I can't just *imagine* that he's *not* in the choir.'

In the second table of the Appendix we give the first 50 items of our single-lemma candidate list with an English translation. For 38% of the items (in bold) it seems to be worthwhile examining their NPI status, or they are already mentioned as NPIs in the literature. For representational reasons we omit presenting the lemma chain candidate list. There, 32% are promising NPI candidates.

unterschätzender	to be underestimated
abbleiben	to get to
beirren	to put off
lumpen	to be stingy
geheuer	scary
nachstehen	to be short of
abreißender	breaking off
unbeträchtlich	unconsiderable
verdenken	to hold it against so
gekannt	known
niet-	rivet (expression *niet- und nagelfest* 'to be nailed down')
langgehen	go along
unähnlich	dissimilar
wegdenken	assume away
dagewesen	precedented
hinnehmbar	tolerable
hingehören	to belong somewhere
verhehlen	to conceal
wegdiskutieren	to rationalize away
totkriegen	to write off sth
überbietender	surmountable
hinwollen	want to go somewhere
wunder	wonder
hinwegtäuschen	to belie something
zimperlich	squeamish
unerheblich	insubstantial
abneigen	to be averse to sth
entblöden	to be shy of doing sth
unterkriegen	to get down
Menschenseele	men's soul
überhören	to overhear
Gegenliebe	approval
spaßen	to trifle with so/sth
missen	miss / do without
behagen	sth. is to so.'s liking

fruchten	to be of any avail
abschätzbar	appreciable
wahrhaben	to believe
hernehmen	to take
draufstehen	to be on the label
auffindbar	detectable
schöner (comp.)	nicer
Hehl	secret
verwinden	to overcome
anbelangen	to concern
deuteln	to argue about
gar	very
Pfifferling	chanterelle (expression *kein Pfifferling wert* 'not worth a straw')
existent	existent
Mördergrube	murderer's pit (expression *(nicht) aus seinem Herzen eine Mördergrube machen* 'to speak frankly')

References

Agirre, Eneko, and Philip, Edmonds, eds. 2006. *Word sense disambiguation. Algorithms and applications*. Dordrecht: Springer.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the 'logicality' of language. *Linguistic Inquiry* 37: 535–590.

Church, Kenneth Ward, and Patrick Hanks. 1990. Word association, norms, mutual information and lexicography. *Computational Linguistics* 1: 22–29.

Condoravdi, Cleo. 2010. NPI licensing in temporal clauses. *Natural Language & Linguistic Theory* 28. doi:10.1007/s11049-010-9115-z

de Swart, Henriëtte. 1998. Licensing of negative polarity items under inverse scope. *Lingua* 105: 175–200.

Drenhaus, Heiner, Douglas Saddy, and Stefan Frisch. 2004. Intrusion effects in the processing of negative polarity items. In *Pre-proceedings of the international conference on linguistic evidence. Empirical, theoretical, and computational perspectives*, eds. Stephan Kepser and Marga Reis, 41–45. Tübingen: Universität Tübingen.

Evert, Stefan. 2005. The statistics of word cooccurrences: Word pairs and collocations. PhD thesis, Universität Stuttgart.

Fauconnier, Gilles. 1975. Pragmatic scales and logical structure. *Linguistic Inquiry* 6: 335–375.

Giannakidou, Anastasia. 1997. The landscape of polarity items. PhD thesis, Rijksuniversiteit Groningen.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency*. Amsterdam-Philadelphia: John Benjamins.

Hoeksema, Jack. 1983. Negative polarity and the comparative. *Natural Language & Linguistic Theory* 1: 403–434.

Hoeksema, Jack. 1997. Corpus study of negative polarity items. In *IV–V Jornades de corpus linguistics*, 1996–1997. http://odur.let.rug.nl/~hoeksema/docs/barcelona.html. Accessed October 2010.

Hoeksema, Jack. 2005. De negatief-polaire uitdrukkingen van het Nederlands. Inleiding en lexicon. Unpublished manuscript, Rijksuniversiteit Groningen.

Hoeksema, Jack. 2010. Dutch *enig*: From nonveridicality to downward entailment. *Natural Language & Linguistic Theory* 28. doi:10.1007/s11049-010-9110-4

Hoeksema, Jack, and Hotze Rullmann, 2001. Scalarity and polarity: A study of scalar adverbs as polarity items. In *Perspectives on negation and polarity items*, eds. Jack Hoeksema, Hotze Rullmann, Víctor Sánchez-Valencia, and Ton van der Wouden, 129–171. Amsterdam: Benjamins.

Kadmon, Nirit, and Fred Landman. 1993. Any. *Linguistics and Philosophy* 16: 353–422.

Klima, Edward. 1964. Negation in English. In *The structure of language*, eds. Jerry A. Fodor and Jerrold J. Katz, 246–323. Prentice Hall: Englewood Cliffs.

Krifka, Manfred. 1995. The semantics and pragmatics of weak and strong polarity items. *Linguistic Analysis* 25: 209–257.

Kürschner, Wilfried. 1983. *Studien zur Negation im Deutschen*. Tübingen: Gunter Narr.

Ladusaw, William. 1980. *Polarity sensitivity as inherent scope relations*. New York: Garland Press.
Lichte, Timm, and Jan-Philipp Soehn. 2007. The retrieval and classification of negative polarity items using statistical profiles. In *Roots: linguistics in search of its evidential base*, eds. Sam Featherston and Wolfgang Sternefeld, 249–266.
Linebarger, Marcia. 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy* 10: 325–387.
Manning, Christopher D., and Heinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge: MIT Press.
Moore, David S., and George P. McCabe. 2006. *Introduction to the practice of statistics*. New York: Freeman.
Rayson, Paul, and Roger Garside. 2000. Comparing corpora using frequency profiling. In *Proceedings of the workshop on comparing corpora,* ACL, 1–8 October 2000, Hong Kong, 1–6.
Saddy, Douglas, Heiner Drenhaus, and Stefan Frisch. 2004. Processing polarity items: Contrastive licensing costs. *Brain and Language* 90: 495–502.
Sailer, Manfred. 2006. *Don't Believe* in underspecified semantics. Neg raising in lexical resource semantics. In *Empirical issues in fromal syntax and semantics*, eds. Olivier Bonami and Patricia Cabredo Hofherr, 375–403.
Vallduví, Enric. 1994. Polarity items, n-words, and minimizers in Catalan and Spanish. *Probus* 6: 263–294.
van der Wouden, Ton. 1992. Beperkingen op het optreden van lexicale elementen. *De Nieuwe Taalgids* 85: 513–538.
van der Wouden, Ton. 1997. *Negative contexts. Collocation, polarity and multiple negation*. London and New York: Routledge.
Vasishth, Shravan, Sven Brüssow, Richard L. Lewis, and Heiner Drenhaus. 2006. Processing polarity: How the ungrammatical intrudes on the grammatical. http://www.ling.uni-potsdam.de/~vasishth/Papers/npi1final.pdf. Accessed October 2010.
von Fintel, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16: 97–148.
Welte, Werner. 1978. *Negationslinguistik. Ansätze zur Beschreibung und Erklärung von Aspekten der Negation im Englischen*. München: Wilhelm Fink Verlag.
Zwarts, Frans. 1995. Nonveridical contexts. *Linguistic Analysis* 25: 286–312.
Zwarts, Frans. 1996. A hierarchy of negative expressions. In *Negation: A notion in focus*, ed. Heinrich Wansing, 169–194. Berlin and New York: de Gruyter.
Zwarts, Frans. 1998. Three types of polarity. In *Plurality and Quantification*, eds. Fritz Hamm and Erhard W. Hinrichs, 177–237. Dordrecht: Kluwer Academic.